



Marek Sroka*
University of Illinois at Urbana-Champaign

COLLABORATION AND COMPETITION IN A DIGITAL AND INTERDISCIPLINARY ENVIRONMENT: HATHITRUST, GOOGLE, OCLC, AND PROJECT GUTENBERG [INTERDYSCYPLINARNA WSPÓŁPRACA I KONKURENCJA W DZIEDZINIE DIGITALIZACJI: HATHITRUST, GOOGLE, OCLC I PROJEKT GUTENBERG]

Abstract: The paper examines in detail the creation of HathiTrust as a collaborative project of sixteen universities and the University of California system to establish a repository for shared digital collections. HathiTrust contains copies of items scanned by Google as well as non-Google content such as scanned books created by Project Gutenberg and the Open Content Alliance, legacy digital content from various institutions, and page images created using publisher-provided PDFs. The author argues that providing access to a huge amount of digital content will require domestic and international collaboration between universities, libraries, and publishers as well as between commercial and non-commercial entities.

DIGITIZATION – HATHITRUST – OCLC-PROJECT GUTENBERG – REPOSITORY

Abstrakt: Współpraca kilkunastu bibliotek amerykańskich doprowadziła do utworzenia cyfrowego repozytorium HathiTrust. HathiTrust zawiera cyfrowe egzemplarze książek stworzonych przez firmę Google i organizację Open Content Alliance. Uzyskanie pełnego dostępu do zasobów cyfrowych wymaga krajowej i międzynarodowej współpracy między bibliotekami, wydawcami, jak również komercyjnymi firmami i organizacjami non-profit.

DIGITALIZACJA – HATHITRUST – OCLC-PROJEKT GUTENBERG - REPOZYTORIUM

*

* *

* Professor MAREK SROKA, Associate Professor of Library Administration and Coordinator of Area Studies Division, University of Illinois Library; MA in English Philology (Jagiellonian University); MA in LIS (University of Illinois Graduate School of Library and Information Science). Two the most important publications: (2009) *The Google Library Project, Internet Archive and Open Content Alliance: Struggle for Universal Access to Digital Collections from the American Perspective*, e-doc. CD [In:] M. Kocójowa ed. (2009). *Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci* [Libraries and Their Clients: Free or Fee Services Supporting Social Communication in Digital Era]. Kraków: Instytut INiB UJ, s. 16–20; (2007) *The Music Collection of the Former Prussian State Library at the Jagiellonian Library in Kraków, Poland: Past, Present, and Future Developments*. "Library Trends" No. 55(3), p. 651–664. E-mail: msroka@illinois.edu

[Prof. MAREK SROKA, Associate Professor of Library Administration and Coordinator of Area Studies Division, University of Illinois Library; absolwent filologii angielskiej (Uniwersytet Jagielloński); MLS (Master of Library Science, University of Illinois Graduate School of Library and Information). Dwie najważniejsze publikacje: (2009) *The Google Library Project, Internet Archive and Open Content Alliance: Struggle for Universal Access to Digital Collections from the American Perspective* [Google Library Project, Internet Archive i Open Content Alliance: walka o powszechny dostęp do kolekcji cyfrowych z perspektywy amerykańskiej], dok. elektr., CD [W:] M. Kocójowa red. (2009). *Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci*. Kraków: Instytut INiB UJ, s. 16–20; (2007) *The Music Collection of the Former Prussian State Library at the Jagiellonian Library in Kraków, Poland: Past, Present, and Future Developments* [Zbiory muzyczne Pruskiej Biblioteki Państwowej w Bibliotece Jagiellońskiej w Krakowie (Polska): przeszłość, teraźniejszość i przyszłość]. "Library Trends" No. 55(3), p. 651–664. E-mail: msroka@illinois.edu]

INTRODUCTION

Many institutions have scanned items in their collections, creating page images and searchable text using OCR. This work has been done on a boutique scale since the 90s. In the last five years the rate of digitization has increased dramatically thanks to support from Google and the Open Content Alliance. As more library collections are digitized there is a growing need to share and archive digitized collections from various institutions. Multi-institutional repositories may play a significant role in providing access to the outputs of various digitization programmes.

HATHITRUST

Launched in October 2008 HathiTrust was established as a collaboration of the thirteen universities of the Committee on Institutional Cooperation and the University of California System. It is a multi-institutional and shared digital repository that provides accessible electronic versions of print titles held by partner institutions. HathiTrust currently has 26 partners, including Columbia University, the University of Chicago, and University of Michigan. The repository contains over 5.6 million currently digitized titles, of which about 15 percent (approximately 864,000 volumes) are in the public domain [HathiTrust, doc. online].

The main goals of HathiTrust include preservation of digital materials of libraries engaged in large-scale digitization as well as providing access to their digital collections. HathiTrust partners were in agreement that "preservation without access is of no value." [York 2009, doc. online, p. 6]. For institutions that have deposited their digital content, HathiTrust is the long-term preservation strategy for that content.

The founders of HathiTrust have been able to overcome many challenges to governance in a variety of complex environments by designing an organizational structure based on two elements: an Executive Committee and a Strategic Advisory Board. The Executive Committee is the decision-making body and consists of university librarians and senior information officers at partner institutions. The main role of the Strategic Board is to develop policies for the repository and its partners.

ACCESS TO HATHITRUST

In 2009 HathiTrust launched a temporary beta catalog. It offers bibliographic searching, including title, author, subject, ISBN/ISSN, publisher, series title, and year of publication. In November 2009, HathiTrust launched a new service allowing for full-text searching capabilities across the repository. The service, based on open source Solr/Lucene technology, makes it possible for users to search public domain and in-copyright works by phrase or keyword.

The repository includes many featured collections that are subject-oriented and listed by a collection name, for example, "Shakespeare," "Polar Bear Expedition," etc.

HATHITRUST AND OCLC (ONLINE COMPUTER LIBRARY CENTER)

Current HathiTrust beta catalog is a temporary feature. A long-term goal is to increase the repository's on-line visibility and accessibility by creating WorldCat (OCLC's Web catalog) records describing its digital content and "linking to the collections via WorldCat.org and WorldCat Local [OCLC, doc. online].

According to John Wilkin, Associate University Librarian, University of Michigan Library and Executive Director of HathiTrust, "The connection between HathiTrust and WorldCat is a natural, WorldCat and HathiTrust are both built by and for libraries, and their pursuit of comprehensiveness will aid our community in pursuit of more effective collection management, as well as integration of services across our institutions" [OCLC, online doc1.].

The collaboration between HathiTrust and OCLC is not only timely but significant as well. One of the biggest challenges facing various digital libraries and repositories is the absence of their holdings and content information in WorldCat-the world's largest bibliographic utility and the world's richest online resource for finding library materials.

In March 2010, OCLC loaded test batches of HathiTrust bibliographic records into WorldCat. OCLC started full-scale loading of HathiTrust bibliographic records after the batches were reviewed by OCLC and the HathiTrust. At the end of March 2010, 1.1 million HathiTrust records were added to WorldCat through OCLC's eContent Synchronization mechanism, and the loading process will continue [HathiTrust, doc. online].

OCLC AND GOOGLE BOOKS LIBRARY PROJECT

HathiTrust is not the only institution partnering with OCLC to provide bibliographic information about their digital collections. Google Books Library Project, which is an effort by Google to digitize collections of major university libraries, will now be represented in OCLC's WorldCat through records of its digitized books. Google sees this collaboration "as part of its mission to make the world's information universally accessible and useful." Jon Orwant, Engineering Manager, Google Books, stated the following reason for the partnership with OCLC: "We've scanned over 12 million books to date, and look forward to the time when every book in the world is discoverable online. Our partnership with OCLC is an important step toward that goal." [OCLC, online doc2.].

WorldCat users will be able to locate digitized books from Google Books Library Project and link to the associated book landing page, and in some cases they will be able to access the full text of eBooks whenever available.

PROJECT GUTENBERG AND MOBILE READER DEVICES

Recently announced alliance between Apple Inc. and Project Gutenberg (the first eBook project that currently has more than 100,000 public domain books) is an example of mobile-izing digital content in the environment where there are two or three times more cell phones than computers. [Gutenberg Project, doc. online]. It is also an example of a collaboration between a big commercial enterprise such as Apple Inc. and one of the first creators of eBooks, namely Project Gutenberg.

Project Gutenberg allows users to download over 30,000 free ebooks to read on their PC, iPhone, iPod, Amazon Kindle, Sony Reader, and recently introduced Apple Inc.'s iPad tablet computer. It also underscores a growing mobile aspect of computing. According to Greg Newby, CEO (Chief Executive Officer) of Project Gutenberg, "the alliance with Apple is not a revenue-generator for his organization, but a way to reach more people." [Wood 2010, p. A-8]. With all new mobile devices having capability to read digital content, including digitized books and ebooks, the access to digital collections will increase and will include many free electronic books.

CONCLUSIONS

Providing access to a huge amount of digital content will require domestic and international collaboration between universities, libraries, and publishers as well as between commercial and non-commercial entities. The first step to provide better information about various digital repositories and their collections requires increased online visibility and accessibility by creating WorlCat (OCLC's Web catalog) records describing digital content of HathiTrust and Google Books Library Project and "linking to the collections via WorldCat.org and WorldCat Local.

Another challenge facing digital libraries is the ever growing number of mobile and portable electronic devices. Recent usability studies of information search on mobile devices seek to understand mobile computing best practice in the design of library services [Hahn 2009]. The mobile revolution is the main reason behind recent partnership and alliance between digital content creators such as Project Gutenberg and computer giants such as Apple Inc., with its recent tablet computer – iPad. As more and more digital content will migrate into mobile and portable devices, there will be even bigger demand for collaboration between major commercial and non-commercial players to provide access to digital collections for both research and entertainment purposes.

REFERENCES

- Gutenberg Project, doc. online. Gutenberg: MobileReader Devices How-To.
http://www.gutenberg.org/wiki/Gutenberg:MobileReader_Devices_How-To [visited 15.04.2010].
- Hahn, J. (2009). On the Remediation of Wikipedia to the iPod. *Reference Services Review* 37(3), p. 272–285.
- HathiTrust. <http://www.hathitrust.org/about> [visited: 13.04.2010].
- HathiTrust, doc. online (2009). Update on March 2010 Activities. http://www.hathitrust.org/updates_March2010 [visited: 14.04.2010].
- OCLC, online doc1. (2009). HathiTrust and OCLC to Work Together to Enhance to Enhance Discovery of Digital Collections. <http://www.oclc.org/us/en/news/releases/20097/htm> [visited: 13.04.2010].
- OCLC, online doc2. (2009). OCLC Adding Records to WorldCat for Google Books Library Project and HathiTrust Digital Library Collections. <http://www.oclc.org/news/releases/2010/201019/htm> [visited: 14.04.2010].
- Wood, P. (2010). Books. *The News-Gazette* (April 3), p. A-8.
- York, J. doc. online (2009). The Library Never Forgets: Preservation, Cooperation, and the Making of HathiTrust Digital Library. <http://www.hathitrust.org/documents/This-Library-Never-Forgets.pdf> [visited: 13.04.2010].